REGULARIZED DECONVOLUTION ON THE CIRCLE AND THE SPHERE

ARNOUD C.M. VAN ROOIJ
Department of Mathematics
Katholieke Univ. Nijmegen
6525 ED Nijmegen
The Netherlands

FRITS H. RUYMGAART
Department of Mathematics
Texas Tech Univ.
Lubbock, TX 79409
U.S.A.

ABSTRACT. In statistical errors in variables models one observes a noisy convolution of an unknown density with a known error distribution, and the problem is to recover the unknown density. This problem amounts to ill-posed inversion of an operator. Such inversion requires regularization. After a brief summary of regularization in general we turn to deconvolution of functions on Abelian groups and focus in particular on the circle. Finally we propose a kind of error in variables model on the sphere (see also Chang (1989)) and mention some ensuing problems.

1. Introduction

In statistics ill-posed problems may arise in situations where we observe or construct a noisy approximation $\hat{g}$ of

$$g = Kf, \tag{1.1}$$

where K is a known transformation of an unknown signal f. The problem is to recover the signal f from $\hat{g}$ which requires inversion of the operator K. This inversion, however, is often ill-posed; even when we restrict to the typical case where K is a one-to-one and continuous linear transformation of a Hilbert space H into itself, K^{-1} need not be defined on all of H (so that $\hat{g}$ may fall outside the range $K(\mathsf{H})$), and K^{-1} need not be continuous. Therefore so-called regularization of the inverse is needed.

The kind of statistical problem to which we will restrict ourselves here is the errors in variables model, where we observe $Y_i = X_i + E_i$, with $X_i \amalg E_i$ and the pairs $(X_1, E_1), ..., (X_n, E_n)$ and hence $Y_1, ..., Y_n$ i.i.d. In this model the unobservable errors E_i have a known density φ, say, the unobservable X_i have unknown density f and the observed $Y_1, ..., Y_n$ have a density g which can be estimated from the data; this density estimator will be denoted by $\hat{g}_n$. In this case the operator K is convolution with φ. Hence to recover f from $\hat{g}_n$ is a deconvolution problem; such problems are ill-posed indeed.

It is clear that in this kind of problem our Hilbert space H is a Hilbert space of functions $L_2(\mathsf{G}, \mathcal{A}, \mu_{\mathsf{G}})$, where G must at least have the structure of a topological group in order that convolution be defined. The symbol μ_{G} denotes an invariant measure on

G. Roussas (ed.), Nonparametric Functional Estimation and Related Topics, 679–690.

G, called Haar measure. For Abelian groups there is no distinction between left and right invariance. If $\oplus$ is the group operation, convolution is defined as

$$(\varphi \circledast f)(x) = \int_{\mathbf{G}} \varphi(x \ominus y) f(y) d\mu_{\mathbf{G}}(y), x \in \mathbf{G}, \tag{1.2}$$

where $x \ominus y = x \oplus y^{-1}$. A theory of regularized deconvolution is indeed possible for certain $\mathbf{G}$, in particular for locally compact Abelian groups.

In the simple examples below of such groups λ will denote Lebesgue measure, restricted or not to a part of $\mathbf{R}$:

(a) $\mathbf{G} = \mathbf{R}$ with $\oplus = +$ and $\mu_{\mathbf{G}} = \lambda$;
(b) $\mathbf{G} = (0, \infty)$ with $\oplus = \times$ and $\mu_{\mathbf{G}} = \mu_+$, the measure with density $(d\mu_+/d\lambda)(t) = t^{-1}, t \in (0, \infty)$;
(c) $\mathbf{G} = \mathbf{R}/\mathbf{Z} = \mathbf{S}$, the unit circle in the plane ($\mathbf{Z}$ the set of all integers); we will make the identification $\mathbf{S} = [0, 1)$, with $\oplus = +$ (modulo 1) and $\mu_{\mathbf{G}} = \mu_0$, Lebesgue measure on $[0, 1)$.

We will focus on $\mathbf{S}$ but briefly mention $\mathbf{R}$ and $(0, \infty)$. Regularized deconvolution on $\mathbf{S}$ will be considered in Section 4.

In higher dimensions the circle generalizes both to the unit sphere $\mathbf{S}^{d-1}$ in $\mathbf{R}^d$ or to $\mathbf{O}(d)$, the orthogonal group in $\mathbf{R}^d$. As we will see below, a modification of Chang's (1989) model leads to a kind of errors in variables model on the sphere, and one can easily imagine how errors in variables on $\mathbf{O}(d)$ arise. However, $\mathbf{S}^{d-1}$ is not even a group (although it is a homogeneous space), and $\mathbf{O}(d)$ is a group but not Abelian. In Section 5 we define an errorsin variables model on the sphere and outline a possible solution. However, we cannot yet present the mathematical details but hope to do so in a forthcoming paper.

Before deconvolution can even be considered we have to face the problem of estimating g from the data $Y_1, ..., Y_n \in \mathbf{G}$. We consider this problem in Section 3 in the case where $\mathbf{G}$ is a Stiefel manifold, which is not in general a group (it is a homogeneous space) but which contains all the examples of present interest (the circle, the sphere, the orthogonal group) as special cases. In Section 2 we briefly discuss a general principle of regularization in Hilbert spaces.

Our references to the vast literature will be rather selective and restricted. An important but not specifically statistical reference in Tikhonov & Arsenin (1977) where the concept of regularization serves the same purpose but is somewhat differently defined. Their technique is applied in Vapnik (1982), often overlooked in the recent statistical literature, to some statistical problems. This author deals with compact operators but deconvolution is not included. For statistical deconvolution on the real line e.g. Carroll & Hall (1988), Fan (1988) and Zhang (1990). Many interesting practical examples can be found in Titterington (1985) and O'Sullivan (1986). In Carroll et al. (1990) a general theory of regularized inversion in Hilbert spaces is reviewed, applied to deconvolution on abstract locally compact Abelian groups, and illustrated with a number of practical statistical problems. However, they don't consider errors in variables on the sphere.

2. Regularized inverses in Hilbert spaces

Let K be an injective operator of a Hilbert space $\mathbf{H}$ into itself; in other words let K be one-to-one and $K \in \mathcal{L}(\mathbf{H})$. By suitable preconditioning (replacing the equation $g = Kf$ by $Tg = TKf$ for suitable one-to-one $T \in \mathcal{L}(\mathbf{H})$) if necessary we may and will assume that K is *strictly positive Hermitian.* It is the contents of the spectral theorem, applied to such operators, that there exists a σ-finite measure space $(\mathcal{X}, \mathcal{A}, \mu)$, a unitary $U : \mathbf{H} \to L_2(\mathcal{X}, \mathcal{A}, \mu)$, a measurable $\gamma : \mathcal{X} \to \sigma(K)$, the spectrum of K, and a multiplication operator $M_\gamma : L_2(\mathcal{X}, \mathcal{A}, \mu) \to L_2(\mathcal{X}, \mathcal{A}, \mu)$, such that

$$K = U^{-1} M_\gamma U, \text{ where } M_\gamma \psi = \gamma \cdot \psi \text{ for } \psi \in L_2(\mathcal{X}, \mathcal{A}, \mu). \tag{2.1}$$

This is well-known for Euclidean spaces where an orthonormal matrix $O \in \mathbf{O}(d)$ exists such that $K = O^t \Lambda O$, with Λ a diagonal matrix. Loosely speaking (2.1) entails that, apart from unitary equivalence, K^{-1} is multiplication by γ^{-1}, an operator that indeed might turn out to be unbounded if $\mathbf{H}$ is infinite dimensional. (In the finite dimensional case $K^{-1} = O^t \Lambda^{-1} O$ which is still continuous but unstable in a sense when the smallest eigenvalue of K (and hence Λ) is close to 0.)

A regularized inverse of K is a sequence of operators $R_m \in \mathcal{L}(\mathbf{H}), m \in \mathbf{N}$, where m is a kind of smoothing parameter, with the property that R_m is close to K^{-1}:

$$\|R_m K f - f\| \to 0, \text{ it as } m \to \infty, \text{ for all } f \in \mathbf{H}. \tag{2.2}$$

If K happens to be invertible (and not just one-to-one) so that $K^{-1} \in \mathcal{L}(\mathbf{H})$, which is the case if $0 < \min\{\lambda : \lambda \in \sigma(K)\}$, we might take $R_m = K^{-1}$ for all $m \in \mathbf{N}$. But even in this case for stability considerations one might prefer to carry out a genuine regularization.

It is possible to construct a regularized inverse once the quantities in (2.1) are given. For $m \in \mathbf{N}$ let us define

$$\rho_m(x) = \begin{cases} \gamma^{-1}(x), & \text{for } x \in \mathcal{X} : \gamma(x) \geq 1/m \\ 0, & \text{elsewhere.} \end{cases} \tag{2.3}$$

It can be shown that (2.2) is satisfied with

$$R_m = U^{-1} M_{\rho_m} U, \text{ where } M_{\rho_m} \psi = \rho_m \cdot \psi \text{ for } \psi \in L_2(\mathcal{X}, \mathcal{A}, \mu). \tag{2.4}$$

This regularization is related to projection. Another interesting regularization is obtained by choosing $\rho_m = (\alpha(m) + \gamma)^{-1}$. In either case we have

$$\|R_m\| \leq m, \quad m \in \mathbb{N}. \tag{2.5}$$

We will use the regularization (2.4).

A proper choice of the regularization parameter m may depend on other parameters in the problem. Let $(\Omega, \mathcal{W}, \mathbf{P})$ be a probability space and the estimators $\widehat{g}_n : \Omega \to \mathbf{H}$ Borel measurable. We will assume that, given $f \in \mathbf{H}$, the element $g = Kf$ can be consistently estimated in the sense that $\mathbf{E}\|\widehat{g}_n - Kf\| = \delta_n(f) \to 0$, as $n \to \infty$, and we know that $\|R_m Kf - f\| = \varepsilon_m(f) \to 0$, as $m \to \infty$. Then it is possible to find for each $n \in \mathbb{N}$ an integer $m(n) \in \mathbb{N}$ (depending on f) such that

$$m(n) \to \infty \quad \& \quad \|R_{m(n)}\|\delta_n(f) \to 0, \ \textit{as } n \to \infty. \tag{2.6}$$

As an estimator for f we propose

$$\widehat{f}_n = R_{m(n)}\widehat{g}_n, \ \textit{with } m(n) \textit{ satisfying (2.6).} \tag{2.7}$$

Since also $\varepsilon_{m(n)}(f) \to 0$, as $n \to \infty$, for such $m(n)$ it follows that

$$\mathbf{E}\|\widehat{f}_n - f\| \leq \|R_{m(n)}\|\delta_n(f) + \varepsilon_{m(n)}(f) \to 0, \ \textit{as } n \to \infty. \tag{2.8}$$

If the rates of $\varepsilon_m(f)$ and $\delta_n(f)$ are known (see (2.5)) for the rate of $\|R_m\|$ it is possible to specify an approximately optimal rate for the r.h.s. in (2.8) by choosing the $m(n)$ in such a way that a balance between the rates of $\|R_{m(n)}\|\delta_n(f)$ and $\varepsilon_{m(n)}(f)$ is obtained.

For at least two important subclasses of operators further specification of the operator U and the function γ in (2.1) is possible: for compact and for convolution operators. By preconditioning Wicksell's problem (estimation of the true radius of a sphere when only radii of random two-dimensional cuts are observed) can be reduced to regularized inversion of a compact operator (Nychka & Cox (1989), Carroll et al. (1990); for recent results by a different approach see e.g., van Es & Hoogendoorn (1988)). Hence this provides a practical example in the first class and, as we have seen, errors in variables models belong to the second class. Here we will restrict ourselves to convolutions.

EXAMPLE 2.1. Let us first briefly consider convolution on $\mathbf{G} = \mathbb{R}$ with $\varphi \in L_1(\mathbb{R}, \mathcal{B}, \lambda)$, symmetric about 0, so that $\mathbf{H} = L_2(\mathbb{R}, \mathcal{B}, \lambda)$. In this case also $L_2(\mathcal{X}, \mathcal{A}, \mu) = L_2(\mathbb{R}, \mathcal{B}, \lambda)$, U is the usual Fourier transform $\mathcal{F}$, and γ the characteristic function of φ denoted by $\tilde{\varphi} = \sqrt{2\pi}\mathcal{F}\varphi$. With all these quantities made explicit the regularized inverse is easily computed. The rate on the right in (2.8) can be improved by restricting f to a Sobolev space; it is still possible to determine U explicitly (Carroll et al.).

EXAMPLE 2.2. Let us next turn to convolution on $\mathbf{G} = (0,\infty)$ with a function $\varphi \in L_1((0,\infty),\mathcal{B},\mu_+)$, symmetric in the sense that $\varphi(x^{-1}) = \varphi(x)$ for $x \in (0,\infty)$. Here $\mathbf{H} = L_2((0,\infty),\mathcal{B},\mu_+)$ and it turns out that $L_2(\mathcal{X},\mathcal{A},\mu) = L_2(\mathbf{R},\mathcal{B},\lambda)$, with $Uf = \mathcal{F}_+ f = \mathcal{F}\ (f(\exp))$ for $f \in L_2((0,\infty),\mathcal{B},\mu_+)$. Finally we have that $\gamma = \tilde{\varphi}_+ = \sqrt{2\pi}\mathcal{F}_+\varphi$. Surprisingly Laplace transforms can be related to this kind of convolution operator. In a statistical context the Laplace transform shows up in a mixture of exponential distributions. This connexion with the Laplace transform is considered in van Rooij & Ruymgaart (1990).

3. Density estimation on Stiefel manifolds

Downs gives an interesting example from vectorcardiography where the sample elements consist of pairs of orthonormal vectors in $\mathbf{R}^3$; such sample points lie in the Stiefel manifold
$\mathbf{S}$ (3,2). The general Stiefel manifold $\mathbf{S}(d,k)$ is defined as the submanifold of $\mathbf{M}(d,k)$, the set of $d \times k$ matrices ($d,k \in \mathbf{N}$ with $1 \leq k \leq d$) which could be identified with $\mathbf{R}^{kd}$, given by

$$\mathbf{S}(d,k) = \{O \in \mathbf{M}(d,k) : O^t O = I_k\}, \tag{3.1}$$

where I_k denotes the $k \times k$ identity matrix. The Euclidean inner product in $\mathbf{M}(d,k)$ satisfies $\langle M_1, M_2\rangle =$ tr $M_1^t M_2$. It is clear that $\mathbf{S}(d,k)$ is a compact subset of $\sqrt{k}\mathbf{S}^{kd-1}$. The Stiefel manifold is not only of interest in its own right but it generalizes both the sphere $\mathbf{S}^{d-1} (k=1)$ and the orthogonal group $\mathbf{O}(d)$ $(k=d)$.

Let $Y_1,\ldots,Y_n$ denote i.i.d. random elements in $\mathbf{S}(d,k)$ with common probability distribution Q. The Haar probability measure on $\mathbf{O}(d)$ induces an invariant probability measure $\overline{\mu}$ on $\mathbf{S}(d,k)$. Let us assume that Q has a continuous positive density g with respect to $\overline{\mu}$. Such a density can be approximated in the following way. Consider the subsets

$$C_O(t) = \{A \in \mathbf{S}(d,k) : \text{tr } A^t O \geq t\}, \quad O \in \mathbf{S}(d,k), t \in [-k,k]. \tag{3.2}$$

We refer to these sets as caps (with center O and height t) because in the special case of the sphere they have such a form. Let us write

$$v(t) = \overline{\mu}(C_O(t)), \ \textit{independent of } O \in \mathbf{S}(d,k), \tag{3.3}$$

due to the invariance of $\overline{\mu}$. It can be shown that

$$P(C_O(t))/v(t) \to g(O), \ \textit{as } t \uparrow k. \tag{3.4}$$

This suggests the estimators

$$\widehat{g}_n(O) = \widehat{P}_n(C_O(t_n))/v(t_n), \textit{ for suitable } t_n \in [-k,k]. \tag{3.5}$$

where $\widehat{P}_n(D)$ is the empirical measure of a Borel set $D \subset \mathbf{S}(d,k)$.

It should be noted that a cap $C_O(t)$ can be produced by intersecting $\mathbf{S}(d,k)$ with a properly chosen closed halfspace in $\mathbf{M}(d,k)$. Let $\mathcal{H}$ denote the class of all such halfspaces. It is well-known that sharp exponential bounds can be obtained for the probabilities on the local fluctuations

$$\mathbf{P}\{\sup_{E \subset \mathcal{H}(D)} |\widehat{P}_n(E) - P(E)| \geq \lambda\}, \lambda \geq 0, \tag{3.6}$$

where $\mathcal{H}(D)$ is the class of subsets obtained by intersecting the measurable $D \subset \mathbf{S}(d,k)$ with all $H \in \mathcal{H}$. With the aid of these bounds the following result can be shown in an almost standard way.

THEOREM 3.1. *Let g be continuous on $\mathbf{S}(d,k)$ with strictly positive minimum, and let $c_n \to \infty$, as $n \to \infty$. Then we have*

$$\left(\frac{nv(t_n)}{c_n|\log v(t_n)|}\right)^{\frac{1}{2}} \sup_{O \in \mathbf{S}(d,k)} |\widehat{g}_n(O) - g_n(O)| \to 0, \textit{ as } n \to \infty, \text{ a.s.}, \tag{3.7}$$

provided that

$$v(t_n) \to 0, \textit{ and } nv(t_n)/\{c_n|\log v(t_n)|\} \to \infty, \textit{ as } n \to \infty, \tag{3.8}$$

where $g_n(O) = \mathbf{E}\widehat{g}_n(O) = P(C_O(t_n))/v(t_n)$.

Comparison of the nonrandom g_n with g follows the usual pattern and combination yields at least the MISE consistency

$$\mathbf{E}\|\widehat{g}_n - g\|^2 = \mathbf{E}\int_{\mathbf{S}(d,k)} \{\widehat{g}_n(O) - g(O)\}^2 d\overline{\mu}(O) \to 0, \textit{ as } n \to \infty, \tag{3.9}$$

as required in Section 2. Condition (3.8) is formulated in terms of the volumes $v(t_n)$. For practical implementation, however, it is the smoothing parameter t_n that we have to choose so that we have to "solve" condition (3.8) for t_n. With the help of Morse's lemma it is possible to convert a rate of $v(t_n)$ into a rate of t_n; see Hendriks et al. (1990) for the details.

4. Regularized deconvolution on the circle

Continuing the examples in Section 2 let us now consider in a little more detail the convolution on $\mathbf{G} = \mathbf{S} = [0,1)$ with a function $\varphi \in L_1([0,1),\mathcal{B},\mu_0)$, symmetric in the sense that $\varphi(\ominus x) = \varphi(x)$ for $x \in [0,1)$ and with $\ominus x = \oplus x^{-1}$ where $\oplus$ is addition modulo 1. Clearly in this example we have $\mathsf{H} = L_2([0,1),\mathcal{B},\mu_0)$. A character on $\mathbf{S}$ is a mapping into the complex unit circle $\mathbf{T}$ which is a continuous group homomorphism. With pointwise multiplication and, in the case of the circle, uniform convergence these characters form a locally compact Abelian group Γ, called the character group of $\mathbf{G}$. Every $k \in \mathbf{Z}$, the set of all integers, determines a character ($i^2 = -1$)

$$\gamma_k(x) = \exp(2\pi i k x), x \in [0,1). \tag{4.1}$$

In this way a topological group isomorphism $\mathbf{Z} \to \Gamma$ is obtained. For simplicity we identify Γ with $\mathbf{Z}$.

We have already seen that in this case $\mu_{\mathbf{G}} = \mu_0$, Lebesgue measure on $[0,1)$, and as an invariant measure on Γ we take $\mu_\Gamma = \#$, the counting measure on $\mathbf{Z}$. Writing $d\mu_0(x) = dx$ we have the following formulas

$$(\varphi \circledast f)(x) = \int_0^1 \varphi(x \ominus y) f(y) dy, x \in [0,1), \tag{4.2}$$

where $\varphi \in L_1([0,1),\mathcal{B},\mu_0), f \in L_1([0,1),\mathcal{B},\mu_0) \cap L_2([0,1),\mathcal{B},\mu_0)$,

$$(\mathcal{F}_0\varphi)(k) = \int_0^1 \varphi(x)\exp(2\pi i k x)dx, \quad (\mathcal{F}_0^{-1}\psi)(x) = \sum\nolimits_{k\in\mathbf{Z}} \psi(k)\exp(-2\pi i k x), \tag{4.3}$$

where $k \in \mathbf{Z}, \varphi \in L_1([0,1),\mathcal{B},\mu_0), x \in [0,1), \psi \in \ell_1(\mathbf{Z})$. We also have

$$\mathcal{F}_0(\varphi \circledast f) = (\mathcal{F}_0\varphi)\cdot(\mathcal{F}_0 f), \textit{ for } \varphi, f \in L_1([0,1),\mathcal{B},\mu_0). \tag{4.4}$$

Because $L_2([0,1),\mathcal{B},\mu_0) \subset L_1([0,1),\mathcal{B},\mu_0)$ it is immediate that $\mathcal{F}_0$ is defined on $L_2([0,1),\mathcal{B},\mu_0)$. It now turns out that $L_2(\mathcal{X},\mathcal{A},\mu) = L_2(\mathbf{Z},\mathcal{P}(\mathbf{Z}),\#) = \ell_2(\mathbf{Z})$ and $U = \mathcal{F}_0$, since it can be shown that $\mathcal{F}_0 : L_2([0,1),\mathcal{B},\mu_0) \to \ell_2(\mathbf{Z})$ and that it is unitary indeed.

Hence, with $Kf = \varphi \circledast f$ for $f \in L_2([0,1),\mathcal{B},\mu_0)$ it follows from (2.1) and the remarks made above that $K = \mathcal{F}_0^{-1}M_\gamma\mathcal{F}_0$, where $\gamma = \mathcal{F}_0\varphi$ and $M_\gamma\psi = \gamma\cdot\psi$ for $\psi \in \ell_2(\mathbf{Z})$. Consequently, a regularized inverse is obtained by taking

$$R_m = \mathcal{F}_0^{-1}M_{\rho_m}\mathcal{F}_0, \textit{ where } M_{\rho_m}\psi = \rho_m\cdot\psi \textit{ for } \psi \in \ell_2(\mathbf{Z}), \tag{4.5}$$

and ρ_m as in (2.3) with $\gamma = \mathcal{F}_0\varphi$ (see also above) and $\mathcal{X} = \mathbb{Z}$. Given any $g \in L_2([0,1), \mathcal{B}, \mu_0)$ we can make this inversion more explicit by putting $R_m g = \mathcal{F}_0^{-1}(\rho_m \cdot \mathcal{F}_0 g) = \xi$. By formal manipulation (that can be justified) we obtain $\mathcal{F}_0\xi = \rho_m \cdot (\mathcal{F}_0 g) = (\mathcal{F}_0\mathcal{F}_0^{-1}\rho_m) \cdot (\mathcal{F}_0 g) = \mathcal{F}_0\big((\mathcal{F}_0^{-1}\rho_m) \circledast g\big)$ according to (4.4), from which we may deduce that $R_m g = (\mathcal{F}_0^{-1}\rho_m) \circledast g$. Because we know that both g and the original f are real functions we should take the real part in (4.3) and replace $\exp(2\pi i k x)$ by $\cos 2\pi k x, x \in [0,1)$. Thus we finally obtain

$$(\mathcal{F}_0^{-1}\rho_m)(x) = \sum\nolimits_{k:(\mathcal{F}_0\varphi)(k) \geq 1/m} \frac{\cos 2\pi k x}{(\mathcal{F}_0\varphi)(k)}, \quad x \in [0,1). \tag{4.6}$$

Summarizing we have proved the following result.

THEOREM 4.1. *Let $Kf = \varphi \circledast f, f \in L_2(\mathbf{S}, \mathcal{B}, \mu_0), \varphi \in L_1(\mathbf{S}, \mathcal{B}, \mu_0)$ with φ symmetric. Suppose that $\widehat{g}_n : \Omega \to L_2(\mathbf{S}, \mathcal{B}, \mu_0)$ are measurable estimators of $g = Kf$ with $\mathbf{E}\|\widehat{g}_n - g\| = \delta_n(f) \to 0$, as $n \to \infty$. Then the estimators*

$$\widehat{f}_n = R_{m(n)}\widehat{g}_n = \big(\mathcal{F}_0^{-1}\rho_{m(n)}\big) \circledast \widehat{g}_n, \quad n \in \mathbb{N}, \tag{4.7}$$

with $\mathcal{F}_0^{-1}\rho_{m(n)}$ given by (4.6), are MISE consistent provided that the $m(n), n \in \mathbb{N}$, satisfy (2.6).

EXAMPLE 4.1. For the practical importance of statistics with values in $\mathbf{S}$ we refer to Mardia (1975). Let us assume that we can observe $Y_1 = X_1 \oplus Z, \cdots, Y_n = X_n \oplus Z_n$, and that the known density of the Z_i is the so-called wrapped standard normal density

$$\varphi_0(x) = \sum\nolimits_{k \in \mathbb{Z}} \varphi(k + x), x \in [0,1), \tag{4.8}$$

where φ is the standard normal density on $\mathbb{R}$. For the computation of (4.5) we note that

$$(\mathcal{F}_0\varphi_0)(k) = \frac{1}{\sqrt{2\pi}} \int_0^1 \{\sum\nolimits_{k \in \mathbb{Z}} \exp(-\frac{1}{2}(x+k)^2 + 2\pi i k x)\} dx = \varphi(2\pi k), \quad k \in \mathbb{Z}. \tag{4.9}$$

This implies that $\{k \in \mathbb{Z} : (\mathcal{F}_0\varphi_0)(k) \geq 1/m\} = \{k \in \mathbb{Z} : |k| \leq a_m\}$, where $a_m = \{(1/2\pi^2)\log(m/\sqrt{2\pi})\}^{\frac{1}{2}}$. In this way we find the estimators

$$\widehat{f}_n = \{\sum\nolimits_{|k| \leq a_{m(n)}} \exp(2(\pi k)^2) \cos(2\pi k(\cdot))\} \circledast \widehat{g}_n(\cdot), \tag{4.10}$$

for suitably chosen $m(n) \to \infty$, as $n \to \infty$.

Often φ will be a smooth density and if it is the density $g = \varphi \circledast f$ will also be smooth. If f is not a priori known to be smooth itself, the problem of recovering such a possibly non-smooth f with the aid of a sample from the smooth g is hard. The difficulty of the inversion is reflected in the speed of convergence. For convolution on $\mathbf{R}$ the dependence of the rate on the number of derivatives is well-known (Carroll & Hall (1988), Fan (1988), Zhang (1990)); in Carroll et al. (1990) the convolution is restricted to a Sobolev space and the operator is again shown to be related with a multiplication but through unitary equivalence with a different Hilbert function space. For convolution on $(0, \infty)$ a similar kind of result is obtained in van Rooij & Ruymgaart (1990). The case of the circle $\mathbf{S}$ is very similar to the real line so that an outline will suffice.

Let us write $f' = f^{(1)}$ for the derivative of f on $\mathbf{S} = [0,1)$ in the sense that $\int_a^b f'(x)dx = f(b) - f(a)$ for all $0 \leq a < b < 1$, and restrict ourselves to the class of smooth functions $\mathbf{H} = \{f \in L_2(\mathbf{S}, \mathcal{B}, \mu_0) : f^{(1)}, \cdots, f^{(p)} \in L^2(\mathbf{S}, \mathcal{B}, \mu_0)\}$, for some $p \in \mathbf{N}$. This class is known to be a Hilbert space, called Sobolev space (Kantorovich & Akilov (1982)) when the inner product

$$\langle f_1, f_2 \rangle_{\mathbf{H}} = \int_0^1 \{f_1(x)f_2(x) + \sum_{j=1}^p f_1^{(j)}(x)f_2^{(j)}(x)\}dx, f_1, f_2 \in \mathbf{H}, \tag{4.11}$$

is used. The restriction of the convolution to $\mathbf{H}$ will be denoted by $K_{\mathbf{H}}f = \varphi \circledast f, f \in \mathbf{H}$. Because $(\varphi \circledast f)' = \varphi \circledast f'$ it is clear that $K_{\mathbf{H}} : \mathbf{H} \to \mathbf{H}$ and is strictly positive Hermitian. For the unitary equivalence we still use the Fourier transform $\mathcal{F}_0$ as defined in (4.3), but with different domain and range viz. $\mathcal{F}_0 : \mathbf{H} \to \ell_{2,\nu}(\mathbf{Z}) = L_2(\mathbf{Z}, \mathcal{P}(\mathbf{Z}), \nu)$, where the measure ν has density

$$(d\nu/d\#)(k) = 1 + \sum_{j=1}^p (2\pi k)^{2j}, k \in \mathbf{Z}, \tag{4.12}$$

with respect to the counting measure $\#$ on $\mathbf{Z}$.

For the isometry note in particular that, for any $\mathcal{F}_0 h = f$ with $h \in \mathbf{H}$ and $f \in \ell_{2,\nu}(\mathbf{Z})$, we have

$$\begin{aligned} \|h\|_{\mathbf{H}}^2 &= \|h\|^2 + \sum_{j=1}^p \|h^{(j)}\|^2 = \|\mathcal{F}_0^{-1} f\|^2 + \sum_{j=1}^p \|(\mathcal{F}_0^{-1} f)^{(j)}\|^2 = \\ &= \|f\|^2 + \sum_{j=1}^p (2\pi k)^{2j} \|\mathcal{F}_0^{-1}((\cdot)^j f(\cdot))\|^2 = \\ &= \|f\|^2 + \sum_{j=1}^p (2\pi k)^{2j} \|(\cdot)^j f(\cdot)\|^2 = \|\mathcal{F}_0 h\|_\nu, \end{aligned} \tag{4.13}$$

where $\|\cdot\|_\nu$ denotes the norm in $\ell_{2,\nu}(\mathbf{Z})$.

Furthermore, in the case of the circle it can be shown that the density estimators $\widehat{g}_n$ of the type discussed in Section 3 satisfy

$$\mathbf{E}\|\widehat{g}_n - g\| = \mathcal{O}(n^{-4/5}), \; as\ n \to \infty, \tag{4.14}$$

assuming the existence and continuity of g'' on $\mathbf{S}$ and provided that we choose $1-t_n = \mathcal{O}(n^{-1/5})$, as $n \to \infty$. This smoothness condition on g will be automatically satisfied since we will assume that φ is smooth.

THEOREM 4.2. *Suppose that $\mathsf{E}\|\widehat{g}_n - g\| = \delta_n(f) = \mathcal{O}(n^{-\alpha})$, as $n \to \infty$, for some $\alpha > 0$ and let $\varphi_0 \in L_1(\mathbf{S}, \mathcal{B}, \mu_0)$ be the wrapped normal density in (4.8). Moreover, let us assume that $f \in \mathbf{H} \subset L_2(\mathbf{S}, \mathcal{B}, \mu_0)$, implying that it is p times differentiable. Then there exist $m(n)$ satisfying (2.6) such that*

$$\mathsf{E}\|\widehat{f}_n - f\| = \mathcal{O}\big((\log n)^{\frac{1}{4}-p}\big), \text{ as } n \to \infty, \tag{4.15}$$

where $\widehat{f}_n$ is given by (4.7).

PROOF. Let us first fix an arbitrary $m \in \mathbf{N}$ and note that, with a_m as in Example 4.1 and exploiting the fact that $\mathcal{F}_0$ is an isometry,

$$\begin{aligned} \varepsilon_m(f) &= \|R_m\varphi \circledast f - f\| = \|\mathcal{F}_0^{-1}(\rho_m \cdot \mathcal{F}_0(\varphi \circledast f)) - f\| = \\ &= \|(\rho_m \cdot \mathcal{F}_0\varphi - 1)\mathcal{F}_0 f\| \leq \\ &\leq \|(\mathbf{1}_{-\infty,-a_m})) \cdot \mathcal{F}_0 f\| + \|(\mathbf{1}_{(a_m,\infty)}) \cdot \mathcal{F}_0 f\|. \end{aligned} \tag{4.16}$$

The two terms in the upper bound can be dealt with in the same way so let us restrict ourselves to the last one.

Using $\|\cdot\|_\nu$ for the norm in $\ell_{2,\nu}(\mathbf{Z})$ as in (4.13), we find

$$\begin{aligned} \|(\mathbf{1}_{(a_m,\infty)}) \cdot \mathcal{F}_0 f\| &= \left\| \left(\frac{\mathbf{1}_{(a_m,\infty)}}{1 + \sum_{j=1}^p (2\pi(\cdot))^{2j}} \right) \cdot (\mathcal{F}_0 f) \cdot \left(1 + \sum\nolimits_{j=1}^p (2\pi(\cdot))^{2j}\right) \right\| = \\ &= \left\langle \frac{\mathbf{1}_{(a_m,\infty)}}{1 + \sum_{j=1}^p (2\pi(\cdot))^{2j}}, (\mathcal{F}_0 f) \right\rangle_\nu \leq \|\mathcal{F}_0 f\|_\nu \cdot \left\| \frac{\mathbf{1}_{(a_m,\infty)}}{1 + \sum_{j=1}^p (2\pi(\cdot))^{2j}} \right\|_\nu = \\ &= \|f\|_{\mathbf{H}} \cdot \left\{ \sum\nolimits_{k > a_m} \frac{1}{1 + \sum_{j=1}^p (2\pi k)^{2j}} \right\}^{\frac{1}{2}} = \mathcal{O}\big((\log m)^{\frac{1}{4}-\frac{1}{2}p}\big), \text{ as } m \to \infty. \end{aligned} \tag{4.17}$$

A proper balance between the values of the quantities on the right in (2.8) in the present case is obtained by choosing $m(n) \sim n^{\alpha/(2\alpha+2)}$, which results in the order on the right in (4.15). Q.E.D.

5. Errors in variables on the sphere

The sphere $\mathbf{S}^{d-1}$ is a homogeneous space under the orthogonal group $\mathbf{O}(d)$, and an element of $\mathbf{O}(d)$ acts on an element of $\mathbf{S}^{d-1}$, where this action will be multiplicatively written as in matrix multiplication. Throughout this section let μ be the invariant probability measure on $\mathbf{O}(d)$ and $\overline{\mu}$ the uniform probability measure on $\mathbf{S}^{d-1}$. All densities will be understood with respect to these measures. In Chang's (1989) spherical regression model an important role is played by random elements on $\mathbf{S}^{d-1}$ with density symmetric about a certain point. For various reasons (e.g. model tests) it might be interesting to describe such a density in terms of the symmetry point and a random element that represents the noise about that point, just like in the case of the real line. It turns out that, given $a \in \mathbf{S}^{d-1}$ and a random element $Z : \Omega \to \mathbf{O}(d)$, the random element $Za : \Omega \to \mathbf{S}^{d-1}$ always has a symmetric density about a. Moreover, all symmetric densities can be generated in this way. See Hendriks & Janssen (1990) where this problem is solved in the more general context of homogeneous spaces.

In a similar way distributions in the spherical regression model with errors in variables (Chang (1989)) can be generated by random elements. Given $X : \Omega \to \mathbf{S}^{d-1}$ and $Z : \Omega \to \mathbf{O}(d)$, one can consider the composition $Y = ZX : \Omega \to \mathbf{S}^{d-1}$. Under the usual assumptions let the density f of X be unknown, the density φ of Z be given, and suppose that we observe $Y_1 = Z_1 X_1, ..., Y_n = Z_n X_n$ with density g. The density g can be estimated e.g. in the manner of Section 3 by $\widehat{g}_n$, say, using the sample elements Y_i. Let us here again focus on the nonparametric problem of recovering f from $\widehat{g}_n$; this is an extension of the errors in variables model on the circle considered in Section 4.

In spite of the fact that the composition is between functions defined on different spaces and $\mathbf{O}(d)$ is not even commutative, the convolution can still very similarly be defined. For $f \in L_2(\mathbf{S}^{d-1}, \overline{\mu})$ and given $\varphi \in L_1(\mathbf{O}(d), \mu)$, both spaces being endowed with their Borel σ-fields, we define the operator

$$g(x) = (Kf)(x) = (\varphi \circledast f)(x) = \int_{\mathbf{O}(d)} \varphi(O) f(O^t x) d\mu(O), \quad x \in \mathbf{S}^{d-1}, \tag{5.1}$$

where obviously $K : L - 2(\mathbf{S}^{d-1}, \overline{\mu}) \to L_2(\mathbf{S}^{d-1}, \overline{\mu})$.

This operator is easily seen to be bounded. The question for which φ the operator is injective needs to be answered. Given an injective operator K, inversion turns out to be ill-posed again. For regularization of the inverse irreducible representations will be an essential tool (Hewitt & Ross (1963)).

Acknowledgement

This work was done while the author was at the Department of Mathematics, Katholieke Universiteit Nijmegen, The Netherlands.

References

1. Carroll, R.J. and Hall, P. (1988) "Optimal rates of convergence for deconvolving a density", J. Amer. Statist. Assoc. 83, 1184-1186.
2. Carroll, R.J., van Rooij, A.C.M. and Ruymgaart, F.H. (1990) "Theoretical aspects of ill-posed problems in statistics", Tech. Report, Kath. Univ. Nijmegen.
3. Chang, T. (1989) "Spherical regression with errors in variables", Ann. Statist. 17, 293-306.
4. Es, A.J. van, and Hoogendoorn, A.W. (1988) "Kernel estimation in Wicksell's corpuscle problem", Biometrika, to appear.
5. Fan, J. (1988) "On the optimal rates of convergence for the nonparametric deconvolution problem", Tech. Report #157, Dept. Statist., Univ. of Cal., Berkeley.
6. Hendriks, H.W.M. and Janssen, J.H.M. (1990) "Densities on homogeneous spaces induced by random elements in the group", Tech. Report, Kath. Univ. Nijmegen, to appear.
7. Hendriks, H.W.M., Janssen, J.H.M., and Ruymgaart, F.H. (1989) "Strong uniform convergence of density estimators on Stiefel manifolds", Tech. Report, Kath. Univ. Nijmegen.
8. Hewitt, E. and Ross, K.A. (1963) Abstract Harmonic Analysis, I, Springer, New York.
9. Mardia, K.V. (1975) Statistics of Directional Data, Acad. Press, New York.
10. Nychka, D.W. and Cox, D.D. (1989) "Convergence rates for regularized solutions of integral equations from discrete noisy data", Ann. Statist. 17, 556-572.
11. O'Sullivan, F. (1986) "A statistical perspective on ill-posed inverse problems", Statist. Sci 4, 502-527.
12. Rooij, A.C.M. van and Ruymgaart, F.H. (1990) "Inversion of noisy Laplace transforms", Tech. Report, Kath. Univ. Nijmegen.
13. Tikhonov, A.N. and Arsenin, V.Y. (1977) Solutions of Ill-Posed Problems, Wiley, New York.
14. Titterington, D.M. (1985) "Common structure of smoothing techniques in statistics", Internat. Statist. Rev. 53, 141-170.
15. Vapnik, V. (1982) Estimation of Dependences Based on Empirical Data, Springer, New York.
16. Zhang, C.H. (1990) "Fourier methods for estimating mixing densities and distributions", Ann. Statist., to appear.